



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
[www.uspto.gov](http://www.uspto.gov)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/748,453	12/26/2000	Mei-Yuh Hwang	M61.12-0339	3442

7590 05/04/2004

WESTMAN, CHAMPLIN & KELLY, P.A.  
Suite 1600 - International Center  
900 Second Avenue South  
Minneapolis, MN 55402-3319

EXAMINER

NOLAN, DANIEL A

ART UNIT	PAPER NUMBER
2654	

DATE MAILED: 05/04/2004

14

Please find below and/or attached an Office communication concerning this application or proceeding.

<b>Office Action Summary</b>	<b>Application No.</b>	<b>Applicant(s)</b>
	09/748,453	HWANG ET AL.
	<b>Examiner</b>	<b>Art Unit</b>
	Daniel A. Nolan	2654

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

#### Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
- Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

#### Status

- 1) Responsive to communication(s) filed on 12 April 2004.
- 2a) This action is FINAL.      2b) This action is non-final.
- 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

#### Disposition of Claims

- 4) Claim(s) 1-21 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) Claim(s) \_\_\_\_\_ is/are allowed.
- 6) Claim(s) 1-21 is/are rejected.
- 7) Claim(s) \_\_\_\_\_ is/are objected to.
- 8) Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

#### Application Papers

- 9) The specification is objected to by the Examiner.
- 10) The drawing(s) filed on \_\_\_\_\_ is/are: a) accepted or b) objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
- 11) The proposed drawing correction filed on 09 June 2003 is: a) approved b) disapproved by the Examiner.  
If approved, corrected drawings are required in reply to this Office action.
- 12) The oath or declaration is objected to by the Examiner.

#### Priority under 35 U.S.C. §§ 119 and 120

- 13) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) All b) Some \* c) None of:  
 1. Certified copies of the priority documents have been received.  
 2. Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.  
 3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- \* See the attached detailed Office action for a list of the certified copies not received.
- 14) Acknowledgment is made of a claim for domestic priority under 35 U.S.C. § 119(e) (to a provisional application).  
a) The translation of the foreign language provisional application has been received.
- 15) Acknowledgment is made of a claim for domestic priority under 35 U.S.C. §§ 120 and/or 121.

#### Attachment(s)

- |   |  |
|---|--|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892)                                    | 4) <input type="checkbox"/> Interview Summary (PTO-413) Paper No(s). _____ . |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)                | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152)  |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449) Paper No(s) <u>2,3</u> . | 6) <input type="checkbox"/> Other: _____ .                                   |

## DETAILED ACTION

1. The text of those sections of Title 35, U.S. Code not included in this action can be found in a prior Office action.

### ***Response to Arguments***

2. Applicant's arguments filed 12 April 2004 have been fully considered but they are not persuasive.

The argument that Smith et al does not make a selection based on scores of speech and acoustic description pertinent to claims 1-4, (page 4 lines 3-4), claim 5 (page 5 lines 5-8), claims 6-8 (page 7 lines 5-9), claims 9-11 (page 8 lines 8-11), claim 18 (page 12 lines 3-5) happens not to be the case, as evinced in the action with reference to the claim (column 12 lines 30-37).

With further regard to claims 6 and 9-11, the argument that Bahl et al does not generate scores for syllable-like units (page 7 line 2 & page 8 lines 12-15) is contradicted with the establishing of *syllables* as a basic unit (column 2 lines 30-34) to be considered with the referenced *phones*. In response to the argument (page 8 last paragraph), the rejection of that feature for claim 9 is maintained by support of the cited reference to Contolini et al (column 6 lines 56-57) with the further disclosure that the *syllabic* characteristic comprises the leaf nodes of the tree and that the score estimator rescores each (column 6 lines 61-65).

With regard to the argument that Gupta et al does not support speech-based phonetic description (page 11 1<sup>st</sup> paragraph), the cited reference does not support the argument stated regarding populating graph (column 5 lines 24-39) and so the rejection is maintained given the understanding that *phoneme rules* generating possible phonetic transcriptions (column 5 lines 25-26).

With regard to claims 19-21, the citation of Schulze is found to apply only to an exception to the invention, those languages that cannot be divided into words or syllable-like units and so the reference is seen to be out of context to the argument.

Regarding claims 20-21, the argument that *there is no latitude in how trigrams are identified in a word* cannot be seen to apply to Schulze given the disclaimer (column 16 lines 37-39 and 56-67).

### ***Claim Rejections - 35 USC § 103***

3. This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

**Smith et al & Häb-Umbach et al**

4. Claims 1–4 are rejected under 35 U.S.C. 103(a) as being unpatentable over Smith et al (U.S. Patent 6,408,271 B) in view of Häb-Umbach et al (U.S. Patent 5,873,061 A).

5. Regarding claim 1, the features employed by Smith et al in *generating phrasal transcriptions for speech recognition dictionaries by permutating word transcriptions for each vocabulary item in an orthographic phrase* read on the features of the *method for adding an acoustic description of a word to a speech recognition lexicon* of the immediate application as follows:

- Smith et al (column 6 lines 15-20) reads on the feature of *converting the text of the word into at least one orthographically derived acoustic description of the word*;
- Smith et al (column 6 lines 42-46) reads on the feature of *generating a score for an orthographically derived acoustic description based in part on a comparison between the orthographically derived acoustic description and a speech signal representing a user's pronunciation of the word*;
- Smith et al (with generating steps 202 & 302 in figures 2 & 3) reads on the feature of *identifying a speech-based acoustic description* 804 in figure 8) *representing the user's pronunciation of the word to produce a speech-based acoustic description of the word and a score for the speech-based acoustic description*; and
- Smith et al (column 12 lines 26-37) reads on the feature of *selecting one of the orthographically derived acoustic description and the speech-based acoustic description as the acoustic description of the word based on the score for the*

*orthographically (column 12 lines 30-31) derived acoustic description and the score for the speech-based acoustic description (column 12 lines 35-37).*

Smith et al does not stipulate that the *acoustic description* be either speech-based or that it is not associated with the text of the word. The invention for *constructing a model of a new word for addition to a word model database of a speech recognition system* of H  b-Umbach et al (figures 1 & 2) reads on the limitation that the *acoustic description is speech-based* (column 3 lines 14-17) and the step of identifying a speech-based acoustic description from the speech signal representing the user's pronunciation of the word, wherein the speech-based acoustic description is not associated with the text of the word (column 4 lines 35-42). It would have been obvious to a person of ordinary skill in the art of speech signal processing at the time of the invention to apply the method/teachings of H  b-Umbach et al to the device/method of Smith et al so as to make recognition not be dependent on transcription by having any subsequent transcription be a separate independent operation apart from of recognition.

6. Regarding claim 2, the claim is set forth with the same limits as claim 1.  
Smith et al (column 13 lines 53-56) reads on the feature of *generating an acoustic model score*.

7. Regarding claim 3, the claim is set forth with the same limits as claim 2.  
Smith et al (column 13 lines 57-60) reads on the feature of *generating an acoustic*

*model score for at least one speech-based acoustic description and using the score as at least part of the score for the speech-based acoustic description.*

8. Regarding claim 4, the claim is set forth with the same limits as claim 3.

Smith et al (802 in figure 8) reads on the feature of *using the same acoustic model* (specified by “*a speech model set*”, column 13 line 52) *to generate both acoustic model scores* (lines 46-56).

**Smith et al, Häb-Umbach et al & Bahl et al<sup>426</sup>**

9. Claim 5 is rejected under 35 U.S.C. 103(a) as being unpatentable over Smith et al in view of Häb-Umbach et al and further in view of Bahl et al<sup>426</sup> (U.S. Patent 5,875,426).

10. Regarding claim 5, the claim is set forth with the same limits as claim 3.

Neither Smith et al nor Häb-Umbach et al teach *generating a language model score*. The Bahl et al<sup>426</sup> method/system for *recognizing speech having word liaisons by adding a phoneme to reference word models* (column 3 lines 55-60) reads on the feature of *generating a language model score for the at least one speech-based acoustic description and* (lines 58-59) *using the language model score as part of the score for the at least one speech-based acoustic description.*

It would have been obvious to a person of ordinary skill in the art of speech signal processing at the time of the invention to apply the method/teachings of Bahl et al

Art Unit: 2654

al'<sup>426</sup> to the device/method of Smith et al & Häb-Umbach et al so as to consider context among the bases of making an acoustic decision.

**Smith et al, Bahl et al'<sup>426</sup> & Bahl et al'<sup>921</sup>**

11. Claims 6–8 are rejected under 35 U.S.C. 103(a) as being unpatentable over Smith et al in view of Häb-Umbach et al and further in view of Bahl et al'<sup>426</sup> and further in view of Bahl et al'<sup>921</sup> (U.S. Patent 6,377,921).

12. Regarding claim 6, the claim is set forth with the same limits as claim 5. Neither Smith et al nor Häb-Umbach et al teach generating a language model score. The Bahl et al'<sup>921</sup> method/system for identifying mismatches between assumed and actual pronunciations of words (column 2 lines 30-38) reads on the feature of generating an acoustic model score and a language model score (lines 39-40 from 18B in figure 1) for a sequence of syllable-like units (column 5 lines 47-48) and (with lines 55-61) the further feature that the speech-based acoustic description is derived from the sequence of syllable-like units.

It would have been obvious to a person of ordinary skill in the art of speech signal processing at the time of the invention to apply the method/teachings of Bahl et al'<sup>921</sup> to the device/method of Smith et al & Häb-Umbach et al to increase precision and avoid prosodic differences by addressing the lower cohesive elements of speech.

13. Regarding claim 7, the claim is set forth with the same limits as claim 6.

Neither Smith et al nor Häb-Umbach et al teach generating a language model score.

Art Unit: 2654

Bahl et al<sup>'921</sup> (with the “phones” of column 6 line 13) reads on the feature of *dividing the sequence of syllable-like units into a sequence of phonemes*, which would have made it obvious to a person of ordinary skill in the art of speech signal processing at the time of the invention to apply the method/teachings of Bahl et al<sup>'921</sup> to the device/method of Smith et al so as to not to over look minor utterances by considering each potential word segment separately.

14. Regarding claim 8, the claim is set forth with the same limits as claim 6.

Neither Smith et al nor Häb-Umbach et al teach *generating a language model score*.

Bahl et al<sup>'426</sup> (column 3 lines 51-53) reads on the feature of *generating a language model score based on a trigram language model for syllable-like units*, which would have made it obvious to a person of ordinary skill in the art of speech signal processing at the time of the invention to apply the method/teachings of Bahl et al<sup>'426</sup> to the device/method of Smith et al & Häb-Umbach et al so as to more quickly isolate candidates from combinations of segments.

**Smith et al, Häb-Umbach et al, Bahl et al<sup>'426</sup> & Contolini et al**

15. Claims 9–11 are rejected under 35 U.S.C. 103(a) as being unpatentable over Smith et al in view of Häb-Umbach et al and further in view of Bahl et al<sup>'426</sup> and further in view of Contolini et al (U.S. Patent 6,233,553 B1).

16. Regarding claim 9, the claim is set forth with the same limits as claim 6.

Neither Smith et al nor Häb-Umbach et al teach *generating a language model score*.

Contolini et al., in the *method and system for automatically determining phonetic transcriptions associated with spelled words*, reads on the feature of generating acoustic model (of claim 4 limiting by claim 1) scores for each of a sequence of phonemes (column 7 line 6) that form the sequence of syllable-like units (column 6 lines 56-57).

It would have been obvious to a person of ordinary skill in the art of speech signal processing at the time of the invention to apply the method/teachings of Contolini et al to the device/method of Smith et al & Häb-Umbach et al so as to be able to relate the results of the recognition that might require correction to those elements that would be familiar to the speaker.

17. Regarding claim 10, the claim is set forth with the same limits as claim 1. Neither Smith et al nor Häb-Umbach et al specify the product reaching a state accessible for human intervention. Contolini et al does so permit such adjustments, with (figure 2) reading on the feature of *displaying a user interface comprising an edit box (item 35) in which a user may enter the text of the word* (as according to the 1<sup>st</sup> lines of the Abstract) *and a list box (item 34) that displays words for which an acoustic description has been previously added to the speech recognition lexicon.*

It would have been obvious to a person of ordinary skill in the art of speech signal processing at the time of the invention to apply the method/teachings of Contolini et al to the device/method of Smith et al & Häb-Umbach et al so as to permit refinements that recognize exceptions to the rules used to set up the vocabulary.

18. Regarding claim 11, the claim is set forth with the same limits as claim 10.

Neither Smith et al nor Häb-Umbach et al specify the product reaching a state accessible for human intervention.

- Contolini et al (figure 2 & column 4 lines 17-25) reads on the features of *receiving an indication that a user has selected a word in the list box* (line 22);
- Contolini et al (column 5 lines 55-56) reads on the features of *retrieving the added acoustic description of the word from the speech recognition lexicon and converting the retrieved acoustic description into an audible signal*.
- It would have been obvious to a person of ordinary skill in the art of speech signal processing at the time of the invention to apply the method/teachings of Contolini et al to the device/method of Smith et al & Häb-Umbach et al so as to audibly confirm the validity of the revision.

#### Gupta et al & Häb-Umbach et al

19. Claim 12–17 are rejected under 35 U.S.C. 103(a) as being unpatentable over Gupta et al (U.S. Patent 6,243,680 B1) in view of Häb-Umbach et al.

20. Regarding claim 12, the *apparatus of Gupta et al for obtaining a transcription of phrases through text and spoken utterances* relates to the features for a *computer-readable medium* of the immediate application as follows:

- Gupta et al (column 1 lines 56-57) reads on the feature of *receiving text of a word* (column 3 lines 27-28) *for which a phonetic description is to be added to a speech*

- recognition lexicon* (line 56) and on the feature of *receiving a representation of a speech signal produced by a person pronouncing the word* (line 57);
- Gupta et al (412 → 400 in figure 4) reads on the feature of *converting the text of the word into a text-based phonetic description of the word*;
  - Gupta et al (402 → 404 in figure 4) reads on the feature of *generating a speech-based phonetic description of the word from the representation of the speech signal*;
  - Gupta et al (406 in figure 4) reads on the feature of *selecting a phonetic description of the word to add to the speech recognition lexicon by selecting between the text-based phonetic description and the speech-based phonetic description based in part on the correspondence between each phonetic description and the representation of the speech signal.*

Gupta et al does not stipulate that the *acoustic description* be either speech-based or that it is not associated with the text of the word. The invention for *constructing a model of a new word for addition to a word model database of a speech recognition system* of Häb-Umbach et al (figures 1 & 2) reads on the limitation that the *acoustic description is speech-based* (column 3 lines 14-17) and the step of identifying a speech-based acoustic description from the speech signal representing the user's pronunciation of the word, wherein the speech-based acoustic description is not associated with the text of the word (column 4 lines 35-42). It would have been obvious to a person of ordinary skill in the art of speech signal processing at the time of the invention to apply the method/teachings of Häb-Umbach et al to the device/method of

Gupta et al so as to make recognition not be dependent on transcription by having any subsequent transcription be a separate independent operation apart from of recognition.

21. Regarding claim 13, the claim is set forth with the same limits as claim 12.

Gupta et al (column 7 lines 29-33) reads on the feature of *generating a plurality of possible phonetic descriptions, using at least one model* (column 4 lines 19-21) to score each possible phonetic description (column 5 lines 3-8) and selecting the possible phonetic description with the highest score as the speech-based phonetic description (column 5 line 16-18).

22. Regarding claim 14, the claim is set forth with the same limits as claim 13.

Gupta et al (column 9 lines 47-62) reads on the feature of *using an acoustic model (of allophones, column 9 line 59) and a language model (using linguistic rules, column 9 line 38)*.

23. Regarding claim 15, the claim is set forth with the same limits as claim 14.

Gupta et al reads on the feature of *using a language model comprises using a language model that is based on syllable-like units* (with the sub-word units of column 9 line 62).

24. Regarding claim 16, the claim is set forth with the same limits as claim 15.

Gupta et al (column 10 lines 6-7) reads on the feature of *generating acoustic model scores for each of the phonemes in a syllable-like unit &* (in column 10 lines 15-18)

*summing the acoustic model scores of the phonemes to generate an acoustic model score for the syllable-like unit.*

25. Regarding claim 17, the claim is set forth with the same limits as claim 12.

- *Gupta et al* (column 10 lines 64-66) reads on the feature of *generating a score for the text-based phonetic description based on the correspondence* (column 11 lines 29-31) *between the text based phonetic description and the representation of the speech signal;*
- *Gupta et al* (column 12 lines 10-16) reads on the feature of *generating a score for the speech-based phonetic description based on the correspondence between the speech-based phonetic description and the representation of the speech signal*
- *Gupta et al* (column 14 lines 24-27) reads on the feature of *selecting the phonetic description with the highest score.*

**Gupta et al, Häb-Umbach et al & Contolini et al**

26. Claim 18 is rejected under 35 U.S.C. 103(a) as being unpatentable over Gupta et al in view of Häb-Umbach et al and further in view of Contolini et al.

27. Regarding claim 18, neither Gupta et al nor Häb-Umbach et al specify a product reaching a state accessible for human intervention, so do not produce audible pronunciations.

- Contolini et al (by selecting the *speaker icon* at the left of figure 2) reads on the feature of *receiving an instruction to generate a audible pronunciation of a phonetic description previously added to the speech recognition lexicon,*
- Contolini et al (column 4 line 52-56) reads on the feature of *retrieving the added phonetic description from the speech recognition lexicon, causing an audible pronunciation to be generated based on the retrieved phonetic description* (column 4 line 40).
- It would have been obvious to a person of ordinary skill in the art of speech signal processing at the time of the invention to apply the method/teachings of Contolini et al to the device/method of Gupta et al & Häb-Umbach et al so as to evaluate generated speech.

#### Schulze & Gupta et al

28. Claims 19-21 are rejected under 35 U.S.C. 103(a) as being anticipated by Schulze (U.S. Patent 6,167,369 A) in view of Gupta et al.

29. Regarding claim 19, the features of the *automatic language identification using both N-gram and word information* of Schulze reads on the *speech recognition system having a language model generated through a process of the immediate application* as follows:

- Schulze does teach the feature of *breaking each word in a dictionary into units* (column 2 lines 55-60) for *n-grams* but does not specify that the units be strictly *syllable-like*. Gupta et al reads on the feature of *breaking each word in a dictionary*

*into syllable-like unit* (with the *sub-word units* of column 9 line 62, with the stipulation given in column 3 lines 53-67). Schulze (column 1 line 29) then teaches the further feature of *for each word* and teaches (by *retaining trigrams* – column 1 lines 24-34 – *into sequences*) *grouping the syllable-like units of the word into n-grams*. It would have been obvious to a person of ordinary skill in the art of speech signal processing at the time of the invention to apply the method/teachings of Gupta et al to the device/method of Schulze so as to separate the contiguous signal into discrete portions corresponding to the dictionary for match processing.

- Schulze (column 12 lines 21-22) reads on the feature of *counting the total number of n-gram occurrences in the dictionary*;
- Schulze (column 12 lines 40-41) reads on the feature of *for each n-gram, counting the number of occurrences of the n-gram in the dictionary and dividing this count by the total number of n-gram occurrences to form a language model probability for the n-gram*.

30. Regarding claim 20, the claim is set forth with the same limits as claim 19. Schulze (column 12 lines 35-37) reads on the feature of *breaking the words by preferring syllable like units that occur more frequently in the dictionary over syllable-like units that occur less frequently*.

31. Regarding claim 21, the claim is set forth with the same limits as claim 20. Schulze (column 12 line 40) reads on the feature of *updating the frequencies of the syllable-like units into which the word is broken*.

***Conclusion***

32. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

33. Any inquiry concerning this communication or earlier communications from the Examiner should be directed to Daniel A. Nolan at telephone (703) 305-1368 whose normal business hours are Mon, Tue, Thu & Fri, from 7 AM to 5 PM.

If attempts to contact the examiner by telephone are unsuccessful, supervisor Richemond Dorvil can be reached at (703)305-9645.

The fax phone number for Technology Center 2600 is (703)872-9314. Label informal and draft communications as "DRAFT" or "PROPOSED", & designate formal communications as "EXPEDITED PROCEDURE". Formal response to this action may be faxed according to the above instructions,

Art Unit: 2654

or mailed to: Mail Stop AF (or CPA, etc. – see Official Gazette, 04 November 2003)  
P.O. Box 1450  
Alexandria, VA 22313-1450

or hand-deliver to: Crystal Park 2,  
2121 Crystal Drive, Arlington, VA,  
Sixth Floor (Receptionist).

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to Technology Center 2600 Customer Service Office at telephone number (703) 306-0377.

Daniel A. Nolan  
Examiner  
Art Unit 2654

DAN/d  
April 25, 2004



RICHEMOND DORVIL  
SUPERVISORY PATENT EXAMINER